An Ontology-Based Retrieval Mechanism for E-learning Systems

Mihaela BRUT
"Alexandru Ioan Cuza" University of Iasi
str.Berthelot nr.16, RO-700483 Iasi
mihaela@infoiasi.ro

***Abstract*—Because the ACM classification system is a standard for the computer science field, we choose it in order to annotate and retrieve the documents and persons involved in this domain, enhancing their mobility inside the e-learning systems. Our retrieval mechanism will consider not only the matches with the concepts involved in a query, but also with their related concepts inside the ACM ontology; the ranking algorithm that we propose will consider the relevance of a concept for a certain document, as well as the annotator competence in the concept topic, and time when the annotations were made.**

***Index Terms*—ontology based retrieval, semantic annotations, ranking algorithm**

I. INTRODUCTION

In order to be easy and suitable retrieved, the materials inside an e-learning platform should be continuously supervised and indexed because their relevance for a certain query depends not only by their content, but by many other factors: the annotations made by the author and by different users; the annotator's competence into the corresponding topic; when was made a certain annotation; the accessing numbers and intervals, as well as the competence of the users which access it.

Taking into discussion the particular case of the Computer Science field and its ACM classification system, we present an ontology-based annotation system together with a search mechanism exploiting the gathered annotations. In order to establish the relevance of each material for a certain topic, a ranking algorithm will be presented.

In the next section we expose the current approaches in ontology-based retrieval mechanisms, discussing various solutions for the ranking algorithm. In the next two sections we present our solution for semantic annotations, respectively for ontology-based retrieval. In the final, we expose our conclusions and future work perspectives.

II. CURRENT APPROACHES IN ONTOLOGY-BASED RETRIEVAL MECHANISMS

The ontology capability of enhancing the data semantics for computer applications lead them to a frequent use in annotations and a more precise information retrieval.

In [9] is proposed a method of searching a document collection via queries which are constituted by ontology concepts, but the ranking algorithm does not consider in fact the relations between ontology concepts.

[10] describes a system with ontology-based annotation and retrieval capabilities for managing the audiovisual information. After the multimedia segmentation process, the annotations are made by specialists, by making reference to some previously selected ontologies, and stored in the semantic base. The search mechanism provide support for semantic queries, based on the some provided search templates.

[5] describes Bibster, a Semantics-Based Bibliographic Peer-to-Peer System, which uses the ACM and SWRC ontologies in order to describe properties of the scientific publications. The retrieval mechanism make also use of a learning ontology, developed on the fly, in order to reflect the actual content of the individual users.

In the system proposed by [6] for annotating the images in order to be easily retrieved, the WordNet ontology is used, for being avoided the ambiguous annotations, as well as for automated extension of image annotations e.g. by synonyms.

[7] presents METIS - a model for media description and classification, that enable user to define customizable media types, metadata attributes, and associations. Media types may be arranged in specialization relationships in order to be able to model taxonomies. The authors provide as case study the implementation of an archive system for research papers and talks in the Computer Science domain, classified according the ACM classification system. The semantic annotations are made through a Web annotation interface. Scientific materials are available for browsing, classification, and annotation through the Web administration interface. The retrieval mechanism is intended to use an implementation of the R-Tree algorithm.

The existing retrieval mechanisms try to implement an efficient ranking algorithm for the results provided for a certain query. Many ranking methods were introduced, based on clever term-based scoring, link analysis, evaluation of user traces etc. [8].

The term-based scoring is the most common ranking method adopted by the IR systems, based on comparing the words (terms) contained in the document and in the query. The ontology-based retrieval systems adapt this method to a concept-based scoring, but there isn't yet an imposed method to this respect.

In the traditional term-base scoring, there is computed the weight of each term that appears in the query, through a statistical analysis. In the case of ontology-based annotation and retrieval mechanism, instead the documents there will be considered their semantic annotations, and the queries will be composed of ontology concepts. To evaluate the relevance of a document for a certain query, there is computed a retrieval status value (rsv). This is equal usually with the weights average, when the query terms are composed through the AND operator, as we consider in our system.

Alongside with average, other possible operators for weights aggregation that are found in the literature are similarity-based evaluation, p-norms, fuzzy logic conjunctive or disjunctive operations [2]. After each document rsv is computed, then documents are ranked on the basis of their rsv.

III. THE ONTOLOGY-BASED ANNOTATION SYSTEM

The proposed ontology-based annotation system consist in a web agent which assist the course editing and annotating operations.

The agent could be used especially inside the e-learning platform supporting the IEEE/LOM standard, because this provide support for semantic extension [1], as could be considered the developed annotations. The integration of a new material into an e-learning platform involves the specification of the material relations with the existing ones. Usually, there is used the Dublin Core / LOM relations set. For enhancing the semantics of the relations between Learning Objects (LOs), the semantics of these relations is refined or modified in [4], and some new relations are introduced.

Our purpose is to relate the LOs with the ACM ontology, and subsequently to find relations between LOs through this ontology. Thus, in order express the rapport between the educational materials and different ontology concepts during the annotation activity we will consider the following relations, in the decreasing importance order:

- *isOnTopic* – for a material which is especially destined to a certain topic
- *hasAsEssentialConcept* – expressing the general keywords for a material
- *usesTheConcept* – expressing the ordinary concepts encountered into material
- *makesReferenceTo* – for marking the referenced concepts from other ACM branch.

1. At the beginning of the editing process, the annotation agent captures the author name, material title, abstract, creation date, storing them in Dublin Core – the most suitable format [3]. Then, it ask the author to select a major topic and some key-concepts from the ACM ontology, in order to be assigned to *isOnTopic* and *hasAsEssentialConcept* relations of the material with the ontology. During the editing process, the annotation agent assist the author to establish matches between the material and ontology concepts, through the *useTheConcept* and *makesReferenceTo* relations.

In case an existing material is accessed by a certain user, the agent capture and include in the material metadata the user name, the beginning and the end of the accessing period. The user could also make some new annotations, adding some new *hasAsEssentialConcept*, *useTheConcept* or *makesReferenceTo* relations.

IV. THE RETRIEVAL MECHANISM

A user who wants to retrieve materials on a certain topic will use a graphical Web interface though which could select the most appropriate ACM ontology concepts. This process could be replaced by an interactive user query reformulation one.

The Web service in charge with the query processing and information retrieval will build, first the query tree, as a sub-tree extracted from the ontology. Then, the query is run over the existing meta information database (developed with the support of the Annotation Agent). For each retrieved document, the score will be calculated, as we mentioned in section 2, as average of the scores obtained by each query's concept, according to its occurrence among document associated metadata.

Because the ACM ontology has 4 levels, we will consider a *level coefficient LC*, with the following values (which have the sum equal 1):

- 0.4 the coefficient for a concept associated with the document which is also one of the query concepts;
- 0.3 the coefficient for a concept situated at one level distance (the concept associated with the retrieved document is parent or child of a query concept);
- 0.2 – the coefficient for 2 levels distance, and
- 0.1 – the coefficient for 3 levels distance;

Also, because the relevance of a concept is provided by its relation with the document, we will consider a Relation Coefficient RC, with the following start values:

- 1 if the concept is referred by *isOnTopic* relation;
- 0.5 if the concept is referred by *hasAsEssentialConcept* relation;
- 0.25 – for *usesTheConcept*;
- 0.12 – for *makesReferenceTo*.

The above listed values are multiplied with the competence coefficient assigned to the author of the metadata which specify this relations, and with the time stamp coefficient (the recent metadata are more valuable then the older ones). The time stamp coefficient is 1 for the last 3 months annotations, and decreases with 0.1 at each 3 months, until 0.1. The author coefficient reflects the author competence into a certain topic, and the corresponding algorithm development is still in progress.

We provide below the proposed ranking algorithm, which assign to each retrieved document receives a score in order to sort them:

```
for each particular retrieved document D
   DS=0;  //The initial document score is 0;
   TN = number of the query concepts; //For average
   for each query concept C
      CS=0; //the concept score is initially 0;
      if (query concept ⊂ document associated metadata)
         then  evaluate RC; // as 1, 0.5, 0.25, 0.12;
               LC = 0.4; CS = CS + LC*RC;
         else  for each CR = concept related to C
               //(situated on a same tree branch as C)
                   evaluate (CR, Q, C); endfor;
      endif;
   DS = DS + CS; endfor;
DS = DS / TN;
```

```
  DS = DS * UC; //UC reflects the effective
document use;
  endfor;
  sort documents according their DS;
  display the sorted document list;
  procedure evaluate (CR, Q, C)
  // contribution of CR to concept C score for
query Q
  LC = 0.1 * (4 - the number of levels between C
and CR);
  if CR among the (query concepts)
       then exit; // CR examined in the main
algorithm
       else if CR among the document metadata
               then evaluate RC for this CR;
                CS = CS + LC*RC; TN = TN + 1;
                    //one concept more in the
final score
            endif;
   endif;.
```

V. CONCLUSIONS AND FURTHER WORK

The ontologies are effectively useful for enhancing the semantic quality of information retrieval process, but the ranking techniques are still into an improvement process. We exposed a retrieval mechanism which evaluate the relevance of a concept for a certain document considering also the related concepts, as well as the time when the annotations were made, and the annotator competence in the concept topic (the evaluation of this competence is still a work in progress, based also on the ACM ontology).

An improvement could be brought to our algorithm by using an incrementally approach: the results for a certain query could be stored, and reused for all the next similar queries, together only with the last created metadata.

REFERENCES

[1] Al-Khalifa H.S., Hugh D. The Evolution of Metadata from Standards to Semantics in E-Learning Applications. Proceedings of Hypertext'06, ACM Press 2006

[2] Boughanem, M., Loiseau, Y., Prade, H., Rank-Ordering Documents According to Their Relevance in Information Retrieval Using Refinements of Ordered-Weighted Aggregations, Adaptive Multimedia Retrieval: User, Context, and Feedback, LNCS, Springer, 2006

[3] Brut, M. Enhancing the Knowledge Management Support within the E-Learning Platforms, Proceedings of DEXA 2007 Workshops, IEEE Press 2007

[4] Engelhardt, M., Hildebrand, A., Lange, D., Schmidt, T.C., Reasoning about eLearning Multimedia Objects. Procedings of SWAMM 06, Springer 2006

[5] Haase, P. Stojanovic, N., Volker, J., Sure, Y. Personalized Information Retrieval in Bibster, a Semantics-Based Bibliographic Peer-to-Peer System, Proc. of I-KNOW ’05, Austria, 2005

[6] Hentschel, C., Nurnberger, A., Schmitt, I., Stober, S., SAFIRE: Towards Standardized Semantic Rich Image Annotation, in Adaptive Multimedia Retrieval: User, Context, and Feedback, LNCS, Volume 4398, 2007

[7] King, R., Popitsch, N., Westermann, U., METIS: a Flexible Foundation for the Unified Management of Multimedia Assets, Multimed Tools Appl no. 33, Springer 2007, 325–349

[8] Long, X., Suel, T., Three-level Caching for Efficient Query Processing in Large Web Search Engines, Proc. of the 14th International Conference on World Wide Web, ACM Press, 2005

[9] Paralic, J., Kostial, I., Ontology-based Information Retrieval. In Proc. of the 14th International Conference on Information and Intelligent systems, IIS 2003, Croatia, 2003.

[10] Tsinaraki, C., Fatourou, E., Christodoulakis, S., An Ontology-Driven Framework for the Management of Semantic Metadata Describing Audiovisual Information, Advanced Information Systems Engineering, LNCS, Volume 2681, Springer Verlag 2003.